

Edge Detection in Hyperspectral Imaging: Multivariate Statistical Approaches

Sergey Verzakov, Pavel Paclík, and Robert P.W. Duin

Information and Communication Theory Group
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands
{s.verzakov, p.paclik, r.p.w.duin}@ewi.tudelft.nl

Abstract. Edge detection is well developed area of image analysis. Many various kinds of techniques were designed for one-channel images. Also, a considerable attention was paid to edge detection in color, multispectral, and hyperspectral images. However, there are still many open issues in edge detection in multichannel images. For example, even the definition of multichannel edge is rather empirical and is not well established. In this paper statistical pattern recognition methodology is used to approach the problem of edge detection by considering image pixels as points in a multidimensional feature space. Appropriate multivariate techniques are used to retrieve information which can be useful for edge detection. The proposed approaches were tested on the real-world data.

1 Introduction

The recent development of sensors makes multichannel images usual objects for analysis. One of the important tools for working with multichannel images is edge detection: finding the places where the properties of image undergo considerable changes. First, detected edges allow the visualization of otherwise difficult to represent multichannel image. Second, it allows to localize objects.

The task of edge detection is connected to the segmentation problem which looks for homogeneous image regions (connected or disconnected). Actually, segmentation answers the question whether the pixel belongs to some segment (cluster, class) and with which confidence. In this sense the results of a segmentation can be used for edge detection: e.g. the pixels with ambiguous confidences can be considered as edges. Another way to employ segmentation for edge detection is to mark pixels as edges whenever the order of confidences (or memberships) are changed [1]. So, the solution of the segmentation task can be easily used for edge detection. On the contrary, having solved the edge detection problem, it is not so straightforward to obtain the segmentation. One can state that two close pixels on the the same side (different sides) of edge belong to the same segment (different segments) but it is much more complicated if possible at all to state this for two arbitrary pixels. So, generally speaking, the task of edge detection provides us with less information but this information is more specific.

For example, in segmentation we need to estimate or get as a prior knowledge the number of segments. This is not needed for the edge detection. One should note that edge detection can be used together with segmentation in order to sharpen edge borders.

The task of detecting edges in gray valued images is very well known. It has a long history and has been thoroughly studied [2,3,4,5]. A review can be found in [6,7]. The same problem for three color, or more general, multichannel images is much less well defined. One of the difficulties in edge detection in multichannel images is the formulation of what is an edge. Indeed, in a gray valued image we can specify the type of intensity profile which we are looking for, i.e. we need to specify a scalar valued function of a scalar argument. (The last statement is not valid for detection of edges in textured images. This task can be converted to edge detection in multichannel images after application of a set of texture detectors.) In a multichannel image we have many more possibilities and it is not always obvious (or it is application dependent) which changes have to be taken into account. The problem of consistent edge definition in multichannel images has not been entirely solved. There are proposals to consider as an overall edge all edges in the separate channels. Hence, possible interaction between channels is neglected. Another approach is to reduce a multichannel image to a gray valued one, e.g., by intensity calculation. It was reported that 90% of the edges detected by this simple approach coincide with edges given by more sophisticated multivariate techniques [8]. However, by this approach we cannot find a change in the color of image which does not involve a change in the intensity level. It implies that channels have to be combined in a non-trivial way: added with different signs or be fused non-linearly [1,9]. But these approaches return again a number of gray valued images. So, the problem of combining is not solved. Very often the question of what is an edge in a multichannel image is not addressed directly but instead gradients of all channels are combined in some way [8,10,11,12].

Other approaches make estimations of the statistical properties of the image in the feature space and learn what can be an edge in this image. Like in [13] where authors proposed to use the "change point" theory for edge detection in gray valued images. The methods employing clustering to extract new channels which are more suitable for the task of edge detection [1,9] can also be considered as such methods, but not completely. They still need to combine the results obtained for different channels. We propose to use the estimation of a joint probability density function (PDF) of two neighbouring pixels. The main idea behind this approach is that pixel combinations typical for edges are rare. So they can be considered as outliers and after learning the joint PDF the edges will be represented by low density regions. One can also think of a modification of this approach which estimates a conditional PDF of the difference between neighbouring pixels. Similar approach was studied in [14], where a complementary cumulative distribution function was used. However, the authors used a distribution modeling technique (cumulative histograms) relevant only for the small number of channels and small number of possible of gray values. We will

use a Parzen density estimation or a mixture of Gaussians. Another important difference between [14] and our work is that we take into account the dependence of distribution on the current pixels location in feature space.

The paper is organized as follows. In the next section we describe some preexisting techniques. Section 3 is devoted to the newly proposed approaches. Then in the section 4 we describe datasets and numerical experiments. The paper is concluded by discussion and conclusions.

2 Preexisting Multichannel Edge Detection Techniques

As we have already mentioned in the introduction, many approaches to edge detection in multichannel images are available. We will review only the most generic ones. All techniques can be split into two large groups. The algorithms of the first type perform image analysis on individual channels and then combine the results (before or after thresholding) without using multivariate statistics in the feature space. However, univariate statistics of gray valued images can still be used for the adaptive selection of the threshold or the size of the filter. We will call this group "Non-statistical or univariate statistical approaches". Another group of algorithms uses multivariate statistics from the beginning and will be referred as "Multivariate statistical approaches". This group can be split into two subgroups. The methods from the first subgroup result again in multichannel images where channels are memberships, confidences, or other types of extracted features. So, combining of the channels is still needed. We will call the methods in this subgroup "Incomplete multivariate statistical approaches". The algorithms from the other subgroup return gray valued images and explicit channel combination is avoided. They are "Complete multivariate statistical approaches".

2.1 Non-statistical or Univariate Statistical Approaches

One of the most popular ways to detect edges in gray valued images is to compute (smoothed) derivatives and then mark as edges all pixels for which the absolute value of the derivative exceeds some threshold and is maximal in some neighbourhood. There are two general ways how to extend this approach to multivariate images. In the first one, spatial partial derivatives are calculated for all channels and combined in some way. For example, 1-norm, 2-norm, or ∞ -norm (max-norm) can be used. In [8] the combination of the gradient magnitudes (instead of its components) is advocated and reported that ∞ -norm combination gives the best results. More sophisticated approach [10,11] suggests to use the largest eigenvalue of the covariance matrix of the set of partial derivatives as an edge magnitude (LEV combination). The result of this combination is the gray valued images of gradient magnitudes. The standard methods of thresholding and edge thinning can be applied to it. Another type of extension performs edge detection for each channel and then combines binary images by, say, the logical OR operation.

Keeping in mind that we are interested in hyperspectral images mostly, we can state that many channels are highly correlated. Thus a very broad spectral

band can obscure more narrow ones during 1–norm, 2–norm, or LEV combinations. The ∞ –norm does not suffer from this. However, it also does not make subband averaging which can lead to a better signal to noise ratio. Another problem, which is encountered by all gradient combination techniques, is that derivatives taken at different channels are scaled differently. So, proper scaling and decorrelation have to be applied to hyperspectral images in order to get combinable gradients.

Another popular method of edge detection in gray valued images employs Laplace of Gaussian (LoG) filters in order to compute smoothed second derivatives. Edges now are defined as zero-crossing points. This approach can be extended to multichannel images in two ways. In the first one LoG is applied to all channels, then the results are summed (maybe with some weights) and thresholding takes place on this image. Note, that this approach is equivalent to the conversion of the image in gray valued image and application of the standard univariate LoG edge detection. Also, it is possible to apply edge detection in each channel and combine the binary results by the OR operator.

The hybrid of the two described ways (maximum of the first derivative and the zero-crossings of the second derivatives) is described in [11,12]. At first, edge magnitudes (contrasts) are calculated as LEV combination of the partial derivatives. Then the zeros of directional derivatives of contrasts are taken as edges.

2.2 Incomplete Multivariate Statistical Approaches

In the multivariate statistical approaches to edge detection one typically employs unsupervised pattern recognition techniques (clustering or density estimation) to use feature space information. Having multichannel images, we may consider each pixel as a point in some feature space. This gives us the possibility to look at the data from the statistical point of view. That is, we can base our algorithm on multidimensional distributions. A few approaches, which use channel statistics were proposed in the past.

The first one [1] consists of fuzzy segmentation of the image and considering zero-crossings of memberships differences: $\Delta_{\mathbf{x}}(i, j) = \alpha_{\mathbf{x}}(i) - \alpha_{\mathbf{x}}(j)$. Here, \mathbf{x} is a pixel position, i, j are cluster indices, and α is a membership. It depends on the task which pairs i, j should be considered: only pair with the largest α or pairs for which $\alpha_{\mathbf{x}}(i)$ and $\alpha_{\mathbf{x}}(j)$ are significant.

The second approach [9] suggests to perform channel extraction based on clustering. Namely,

$$J_{\mathbf{x}}(i, j) = \frac{(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T}{\|\boldsymbol{\mu}_i - \boldsymbol{\mu}_j\|} \mathbf{I}_{\mathbf{x}}$$

Here $\boldsymbol{\mu}_i$ is the centroid of i -th cluster. Thus, we get $N(N - 1)/2$ new channels. Authors suggest to combine edge magnitudes of these channels by ∞ –norm and then apply thresholding. The computational cost can be decreased by taking into account only a few neighbouring clusters at each pixel position.

2.3 Complete Multivariate Statistical Approaches

The technique proposed in [14] involves computation of cumulative multidimensional histogram of pixel differences. In that way, a new distance between pixels is introduced. This approach is appropriate only for images with a small number of channels and not a large number of gray levels. In the next section we propose to use density estimators which are more suitable for high dimensional data.

3 Proposed Statistical Techniques

3.1 Joint Probability Density Functions of Neighbouring Pixels

The main hypothesis which will be used for developing the technique is that edges are pretty rare events. This is a natural assumption for many real-world images. Thus, one can conclude that a pair of neighbouring pixels positioned on different sides of an edge (or pair in which one of the pixels is pure and another is the transitional one, or both are transitional) should be also rare compared to pixel pairs in the interior regions.

Let us define by $\mathbf{I}_{\mathbf{x}}$ a d -dimensional vector of channel intensities of a multi-channel image \mathbf{I} at some pixel position $\mathbf{x} \in \mathbb{R}^2$. Further, suppose that a $\mathcal{N} \subset \mathbb{R}^2$ is a set of local shifts. Then the pixel $\mathbf{y} = \mathbf{x} + \mathbf{r}$ is the neighbouring pixel of the pixel \mathbf{x} . The joint PDF $\rho(\mathbf{I}_{\mathbf{x}}, \mathbf{I}_{\mathbf{x}+\mathbf{r}})$ has to be small if \mathbf{x} is situated at an edge orthogonal (or at least not collinear) to \mathbf{r} .

The proposed approach consists of the estimation of $\rho(\mathbf{I}_{\mathbf{x}}, \mathbf{I}_{\mathbf{x}+\mathbf{r}})$, $\mathbf{r} \in \mathcal{N}$ by an appropriate technique like Parzen or Mixture of Gaussians density estimation. An estimated PDF can be used for edge detection. For the edge direction (i.e. direction along which the edge is locally extended) perpendicular to the shift \mathbf{r} it gives a gray valued image of edge magnitudes $m_{\mathbf{x},\mathbf{r}}$ which is calculated as

$$m_{\mathbf{x},\mathbf{r}} = 1 - \rho(\mathbf{I}_{\mathbf{x}}, \mathbf{I}_{\mathbf{x}+\mathbf{r}}) / R_{\mathbf{r}}$$

$$R_{\mathbf{r}} = \max_{\mathbf{x}} \rho(\mathbf{I}_{\mathbf{x}}, \mathbf{I}_{\mathbf{x}+\mathbf{r}})$$

To detect edges independently of their directions one may combine directional magnitudes as $m_{\mathbf{x}} = \max_{\mathbf{r} \in \mathcal{N}} m_{\mathbf{x},\mathbf{r}}$.

The binarization of this image can be done by putting threshold at some suitable percentile. To get more thin edges one can consider non-maximum suppression techniques similar to the ones used in gray valued case [7]. One can also think about smoothing obtained edge magnitudes. This is expected to be useful in noisy images.

Another problem is caused by the large dimensionality of the data. We need to estimate a PDF in the doubled feature space ($2d$). A proper dimensionality reduction technique, like PCA, can be used to solve this issue. It seems more reasonable to perform dimensionality reduction in doubled (not original) feature spaces.

3.2 Conditional Probability Density Functions of Neighbouring Pixels Difference

Having defined by $\delta\mathbf{I}_{\mathbf{x},\mathbf{r}} = \mathbf{I}_{\mathbf{x}+\mathbf{r}} - \mathbf{I}_{\mathbf{x}}$ the difference between neighbouring pixels, the joint distribution can be rewritten as $\rho(\mathbf{I}_{\mathbf{x}}, \mathbf{I}_{\mathbf{x}+\mathbf{r}}) \equiv \rho(\mathbf{I}_{\mathbf{x}}, \delta\mathbf{I}_{\mathbf{x},\mathbf{r}})$. So, it is

possible to reconsider the above described method as a search for rare combinations of a pixel and a difference. But one can argue that some types of pixels are represented much less often than others. Because of this their pairs are also rare, although they do not represent any edge. To rule out such an unwanted situation, we propose to use conditional PDFs: $\rho(\delta\mathbf{I}_{\mathbf{x},\mathbf{r}}|\mathbf{I}_{\mathbf{x}}) = \rho(\mathbf{I}_{\mathbf{x}}, \mathbf{I}_{\mathbf{x}+\mathbf{r}})/\rho(\mathbf{I}_{\mathbf{x}})$. Then edge magnitudes are defined as

$$\begin{aligned} m_{\mathbf{x},\mathbf{r}}^c &= 1 - \rho(\delta\mathbf{I}_{\mathbf{x},\mathbf{r}}|\mathbf{I}_{\mathbf{x}})/R_{\mathbf{r}}^c \\ R_{\mathbf{r}}^c &= \max_{\mathbf{x}} \rho(\delta\mathbf{I}_{\mathbf{x},\mathbf{r}}|\mathbf{I}_{\mathbf{x}}) \\ m_{\mathbf{x}}^c &= \max_{\mathbf{r} \in \mathcal{N}} m_{\mathbf{x},\mathbf{r}}^c \end{aligned}$$

Consequently, only differences $\delta\mathbf{I}_{\mathbf{x},\mathbf{r}}$ which are rare for the pixel values $\mathbf{I}_{\mathbf{x}}$ will be considered as edges. Note, that to make a consistent conditional density estimation, the dimensionality reduction has to be done in original feature space.

4 Experimental Study

4.1 Datasets

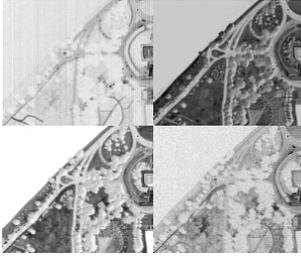
To make a comparison between the discussed techniques a number of datasets have been used. The first one is the hyperspectral image of Washington DC Mall from [15]. This is a 191-channel airborne hyperspectral image of size 1280-by-307. The sensor system used in this case measured a response in 0.4 to 2.4 μm region of the visible and infrared spectrum. The task of edge detection can be formulated as a contour detection of homogeneous areas (roofs, roads, paths, trees, grass, water and shadows). The image itself is too large to be handled at once and we split it into 20 smaller 128-by-153 images. We have used only the upper left one (DC_{1,1}). This is a "busy" image with many details and large number of channels. It is expected that the multivariate statistical approach will be more suitable for reliable edge detection than adapted gray valued image analysis techniques.

Another image from [15] is a 12-channel (0.4 to 1 μm) 949-by-220 airborne image. We have split this image into 3 images of sizes 316-by-220 and used the middle one (FLC₁₂). This image contains much simpler scene and has moderate number of channels. Thus, the usage of adapted image analysis techniques is expected to be enough.

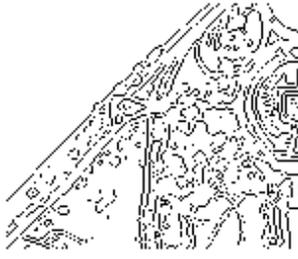
The third collection of images contains 5 microscopic SEM/EDX 8-channel 128-by-128 images of chemical substances [16] from which only CHM₂ has been used. Image is extremely noisy both in the spectral and spatial domains.

4.2 Experiments

We have conducted a set of experiments on the above described images. The results are presented on Fig. 1-3. The first subfigure of each figure shows four typical channel images of multichannel image. The second subfigure represents the edges detected by non-statistical approach. Actually, for all datasets we



(a) Dataset

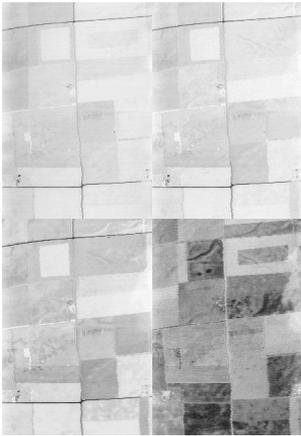


(b) Sobel method



(c) proposed technique

Fig. 1. $DC_{1,1}$



(a) Dataset

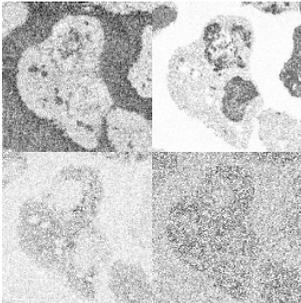


(b) Sobel method



(c) proposed technique

Fig. 2. $FLC_{1,2}$



(a) Dataset



(b) Sobel method



(c) proposed technique

Fig. 3. CHM_2

computed channel gradients with the help of Sobel operator and combined them by ∞ -norm. Binarization and non-maxima suppression were performed as it is suggested in [7]. Zero-crossing methods or other channel gradients combination rules give similar or worse results. The rightmost subfigures show the results of the proposed technique. On all images the PCA dimensionality reduction was performed with preserving 95% of total variance. EM-algorithm was used to estimate PDFs as Mixture of Gaussians with 10 components. This number of components proved to be reasonable for all images. The the results obtained by Parzen density estimator are similar to the presented ones. Only the CHM₂ edge magnitudes were smoothed by Gaussian filter with window size 17-by-17 and $\sigma = 2.67$, because other images did not benefit from smoothing.

5 Discussion and Conclusions

In this paper the task of edge detection in multichannel and especially in hyperspectral images was studied. The goal was not to propose a fast real-time algorithm but to try to develop a consistent approach to edge detection for multichannel images. The new approach based on the statistical pattern recognition was proposed. Instead of explicit definition of the edge we try to learn it by looking for improbable pixel combinations. The comparison of the results of conventional methods and proposed one shows that for high-dimensional complicated images detected edges are very similar (Fig. 1). For simpler images the image processing approach gives the better result (Fig. 2). For noisy chemical data our approach allows to obtain closed thin contours. The proposed approach is computationally expensive. So, it is necessary to develop faster density approximation algorithms. Another possible topic for the future research is incorporation spatial relations of pixels into the density estimation.

Acknowledgments

Authors would like to thank J. von Frese, A. Harol, P. Juszczak, M. Loog, and E. Pekalska for the helpful brainstorming. This research was supported by the Technology Foundation STW, applied science division of NWO and the technology program of the Ministry of Economic Affairs.

References

1. T.L. Huntsberger and M.F. Descalzi. Color edge detection. *Pattern Recognition Letters*, 3:205–209, 1985.
2. D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London*, B207:187–217, 1980.
3. R. Nevatia and K. R. Babu. Linear feature extraction and description. *Computer Graphics Image Processing*, 13(3):257–269, July 1980.
4. R. M. Haralick. Digital step edges from zero crossing of second directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1):58–68, 1984.

5. J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.
6. T. Y. Young and K.-S. Fu, editors. *Handbook of pattern recognition and image processing*. Academic Press, Inc., 1986.
7. J. S. Lim. *Two-dimensional signal and image processing*. Prentice-Hall, Inc., 1990.
8. T. Kanade and S. Shafer. Image understanding research at cmu. In *Proceedings of an Image Understanding Workshop*, volume I, pages 32–40, 1987.
9. H. Tao and T. S. Huang. Color image edge detection using cluster analysis. In *Proceedings of the 1997 IEEE International Conference on Image Processing (ICIP 1997)*, volume 1, pages 834–837, 1997.
10. S. Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, 33(1):116–125, 1986.
11. A. Cumani. Edge detection in multispectral images. *CVGIP: Graphical models and Image Processing*, 53(1):40–51, 1991.
12. W. Alshatti and P. Lambert. Using eigenvectors of a vector field for deriving a second directional derivative operator for color images. In *Proceedings of the 5th International Computer Analysis of Images and Patterns Conference (CAIP 93)*, volume 719 of *Lecture Notes in Computer Science*, pages 149–156, 1993.
13. J. S. Huang and D. H. Tseng. Statistical theory of edge detection. *Computer Vision, Graphics, and Image Processing*, 43(3):337–346, 1988.
14. M. Pietikainen and D. Harwood. Edge information in color images based on histograms of differences. In *Proceedings of The 8th International Conference on Pattern Recognition Conference (ICPR 86)*, volume 1, pages 594–596, 1986.
15. D. Landgrebe. *Signal theory methods in multispectral remote sensing*. John Wiley & Sons, 2003.
16. Pavel Paclík, R. P. W. Duin, G. M. P. van Kempen, and R. Kohlus. Segmentation of multi-spectral images using the combined classifier approach. *Image Vision Comput.*, 21(6):473–482, 2003.